# FINAL REPORT: September 3, 2010

**DISTRIBUTION STATEMENT A.: Approved for public release; distribution is unlimited**

**Cooperative Agreement**

W911NF-09-2-0023 U. S. Army Research Laboratory

**Project Title**

Mental Workload Manipulation Using Multiple Homogeneous Tasks: Performance Effects

**Principal Investigator**

Herbert A. Colle
Department of Psychology
Wright State University
Dayton, OH 45435
Phone: 937-775-2391 email: collewsu@yahoo.com



**DISTRIBUTION LIST:**

**ARL Cooperative Agreement Manager (CAM)**

Dr. Daniel Cassenti
U.S. Army Research Laboratory
Attn: AMSRD-ARL-HR-SE
Aberdeen Proving Ground, MD 21005
Phone: 410-278-5859, email: daniel.cassenti@us.army.mil

**Agreements Administration Office**

Office Of Naval Research (ONRRO) Chicago
230 South Dearborn
Room 380
Chicago, IL 60604-1595
Phone: 312-886-5423, Fax: 312-353-2094, email: ONR_Chicago@onr.navy.mil

**Grants Officer**

Kathryn McManus
U.S. Army Contracting Center
ATTN: CCRD-RT
4300 S. Miami Blvd.
Durham, NC 27703
Phone: 919-549-4264, Fax: 919-549-4388, email: kathryn.mcmanus@us.army.mil

**Defense Technical Information Center (DTIC)**

8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218

## Technical Report

The research project was a 12 month endeavor to conduct empirical research, the results of which provide a better understanding of multi-tasking and mental workload, concepts that are increasingly important to Army research. Soldiers face increasing mental workload demands as they face tasks created by new personnel demands and technology demands. This research provides a deeper understanding of how performance declines as mental workload demand increases.

The theoretical underpinnings of this research include (a) multiple resource theory (Wickens, 2002) and (b) the accomplishment model of average mental workload (Colle & Reid, 1988, 1997, 1999, 2005). (a) Multiple resource theory assumes that multiple dimensions are needed to describe an operator's mental workload with separate capacities for each dimension. These capacities limit task performance, but the theory does not specify a performance-resource function except that they are monotonically related. Multiple resource theory assumes that there are three mental workload dimensions: stages (cognitive, response), processing codes (verbal, spatial), and sensory modalities (auditory, visual). (b) The accomplishment model is an example of a fundamental measurement theory (Krantz, Luce, Suppes, & Tversky, 1971).) The accomplishment model assumes that mental workload is an extensive measurement system so that its methods are used to scale task performance in order to recover underlying mental workload levels (a performance-resource relationship). Both approaches are applicable if testing is conducted using sets of tasks that are homogeneous, so that the same resource dimension is needed to perform all tasks. For homogeneous tasks, theory and data agree that mental workload increases as the number of tasks being performed concurrently increases. This is a basic assumption of the C3TRACE (Command, Control, and Communications: Techniques for the Reliable Assessment of Concept Execution, Kilduff, Swoboda, & Barnette, 2005) and IMPRINT (IMproved Performance Research INTegration Tool; Archer, 1998) human behavior modeling systems. The most homogeneous set of tasks is created when multiple instances of the same task must be performed concurrently.

The research is unique because the number of concurrently performed tasks (from 1 to 8) was used to determine the total mental workload. Typically, mental workload has been manipulated by changing the levels of a task or by concurrently performing different tasks (typically 2) to manipulate workload. These previous manipulations were based on the experimenter's qualitative judgment of mental workload levels. In contrast, the present approach was a quantitative manipulation that lends itself to quantitative assessments of performance changes.

One to eight instances of the same task were performed concurrently. Each additional homogeneous task should have added a constant amount of underlying mental workload.

The experimental task investigated performance on a scanning/visual search task. The display for each task consisted of 4 letters that were presented in a vertical column on a display monitor. A participant's task was to find a target letter if it was present. Each display task had either one instance of the target letter or no instance of it. The target was the letter "X." A column was presented for a fixed period of time, the display duration. Participants were required to press a key to indicate that there was a target present before the display ended. At the end of the display duration, a new display was presented. Successive displays were presented until the test trial ended. When more than one task was presented concurrently, each task display was another column of letters and the response to it was a different key press. Columns and the target-present keys were spatially compatible. The leftmost display column used the leftmost response key and the rightmost display column used the rightmost response key with intermediate keys correspondingly mapped. The onset of all concurrent tasks was synchronized so that the display onsets and offsets coincided.

The proposed research manipulated three different factors of the scanning/search visual task. One factor was the number of concurrently performed tasks (CPTs). This factor was the major manipulation. Number of CPTs varied from 1 to 8 so that performance as a function of number of CPTs could be observed. The other two factors were display duration and pause lag between display screens. These two factors each had three levels. Display duration controlled the rate at which successive displays were presented. Display duration/rate was important because Cassenti and Meyer (2008) found that the effect of the number of CPTs also depended on the effective display duration for visually tracking performance. After each display duration ended, the next display either started immediately (zero pause lag) or a fixed duration pause lag occurred before the next display screen was presented. During the pause lag, the display was blank. Thorne (2006) proposed equations to compensate for this dead time. The pause factor allowed an assessment to see if they produced performance that is equivalent to the no pause condition. Manipulating both pause lag duration and display duration contributed to our understanding of how performance curves can be implemented in the IMPRINT network model.

The experimental design was a 3 x 3 x 8 mixed factorial design. The between-subject factor was pause lag duration. The two repeated-measures factors were display duration and number of CPTs. Participants were randomly assigned to each of the three levels of the between-subjects factor of pause lag duration, 24 participants to each level. Levels of display duration and number of CPTs were balanced across participants using balanced Latin squares. This design allowed the generation of 9 different curves relating performance to number of CPTs. It allowed a direct comparison of these 9 performance curves. Pilot studies were conducted to determine the best levels of display duration (1000, 2000, and 3000ms), pause lag (0, 250, and 500ms) and other parameters such as target probability (p = .25) and the number of displays per trial (40). All participants were tested individually. Researchers at ARL provided software programming using the E-Prime programming language, which was used with some modification to test participants.

The second objective of obtaining subjective ratings of mental workload was studied by using the Subjective Workload Assessment Technique, SWAT (Reid & Nygren, 1988). SWAT ratings were collected from each participant after each test trial. Thus, SWAT ratings were obtained for each of the cells of the 3 x 3 x 8 mixed factorial design. All participants participated in card sort scaling in small groups of 12 or less prior to the mental workload performance test session. Scaling information was used to convert the ratings on the three SWAT dimensions to and overall mental workload scale of 0 to 100. Reid and Colle (1988) proposed a mental workload redline for SWAT ratings. The ratings from the research were included to test this SWAT redline rule.

The research study was conducted at Wright State University (WSU) after being approved by its IRB. The head of the internal review board (IRB) of the U.S. Army research Laboratory (ARL) has found that the University's IRB and not the ARL IRB may review and approve the study. No involvement of ARL personnel was included in conducting the studies.

After the research testing was completed, researchers at both ARL and the University analyzed statistics, interpreted data, and shared their findings with each other. Besides the analyzed statistics, all raw data collected at WSU were sent to Dr. Cassenti at ARL. He analyzed the data in the same way that Cassenti and Kelley (2006, 2008) analyzed similar studies. Each measure was plotted by the number of tasks performed at once. Regression methods were employed to find the best fitting curve and the knot and splines method was used, which found some distinct segment trends composing the overall trends. The data were consistent with Drs. Cassenti and Kelley's theoretical approach, which assumes that performance curves have multiple segments separated by the spline curve knots. The curves followed the expected shallow, steep, shallow decline pattern as mental workload increased with increasing number of tasks. In addition, each study was analyzed using accomplishment scores for a fixed time interval, as they were in Colle and Reid (2005) and in Calkin (2007). These accomplishment curves also were related to the SWAT ratings. Performance and subjective mental workload redlines were estimated, but there were estimation difficulties because of the range of performance values that were generated. WSU and ARL researchers jointly decided that the results were publishable and are working on a joint manuscript, which will be submitted in 2011. Because the results were most relevant to the theories of Cassenti and Kelley, it was decided that they would have first authorship and that they would take the lead on writing the manuscript.

References


Archer, S. (1998). *Improved Performance Research Integration Tool (IMPRINT)Analysis Guide 4.0.* Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Calkin, B. (2007). Parameters affecting mental workload and the number of simulated UCAVs that can be effectively supervised. M. S. Thesis, Wright State University.

Cassenti, D. N., & Kelley, T. D. (2006) Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1147-1152). Santa Monica, CA: Human Factors and Ergonomics Society.

Cassenti, D. N., & Kelley, T. D. (2008). Towards a workload-performance prediction tool. Proceedings of the 30th Annual Meeting of the Cognitive Science Society, New York.

Colle, H. A., Amell, J. R., Ewry, M. E., & Jenkins, M. L. (1988). Capacity equivalence curves: A double trade-off curve method for equating task performance. *Human Factors, 30,* 645–656.

Colle, H. A., & Reid, G. B. (1997). A framework for mental workload research and applications using formal measurement theory. *International Journal of Cognitive Ergonomics, 1,* 303–313.

Colle, H. A., & Reid, G. B. (1999). Double trade-off curves with different cognitive processing combinations: Testing the cancellation axiom of mental workload measurement theory. *Human Factors, 41,* 35–50.

Colle, H. A., & Reid, G. B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. *International Journal of Aviation Psychology, 15*,303-319.

Kilduff, P. W., Swoboda, J. C, & Barnette, D. D. (2005). *Command, Control, and Communications: Techniques for the Reliable Assessment of Concept Execution (C3TRACE) Modeling Environment: The Tool.* Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). Foundations of measurement (Vol. 1). New York: Academic.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT values for predicting operator overload. In *Proceedings of the Human Factors Society 32nd annual meeting* (pp. 1414–1418). Santa Monica, CA: Human Factors and Ergonomics Society.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: Elsevier.

Thorne, D. R. (2006). Throughput: A simple performance index with desirable characteristics. *Behavior Research Methods, 38*, 569-573.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science, 3(2)*, 159-177.

## Business Status Report

**Inventions**

No subject inventions were developed or utilized during the 12 month period of the grant.

**Deliverable**

The primary deliverable was the raw data and statistically analyzed results and descriptive figures, which were provided to Dr. Cassenti, the ARL CAM.

**Joint Papers and Presentations/Journal Article**

We established a good cooperative working relationship, which continues informally after the end of the granting period. This includes working on a jointly authored manuscript, which is planned to be submitted for publication in 2011. The lead author of the manuscript will be Dr. Cassenti.

## Final Direct Expenditures

The major expenditures were for the graduate research assistant, Ms. Elizabeth McGregor, and undergraduate student research assistants. In addition, some funding was provided for materials and supplies needed to conduct the experiments. These expenditures were consistent with the expenditures budgeted in the Cooperative Agreement.

**Final Direct Expenditures were**

| | |
|---|---|
| Graduate Research Assistantship stipends | $8,528.71 |
| Undergraduate Research Assistants | $4,962.08 |
| Materials and Supplies, etc. | $1,313.01 |
| Total Direct Expenditures | $14,803.80 |